(4)

USAISEC

US Army Information Systems Engineering Command
Fort Huachuca, AZ 85613-5300

U.S. ARMY INSTITUTE FOR RESEARCH
IN MANAGEMENT INFORMATION,
COMMUNICATIONS, AND COMPUTER SCIENCES
(AIRMICS)

EXPERT SYSTEMS IN EVALUATION METHODOLOGY

(ASQBG-A-89-034)

July, 1989

AIRMICS
115 O'Keefe Building
Georgia Institute of Technology
Atlanta, GA 30332-0800

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704--188
Exp. Date: Jun 30, 1986

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS
NONE

2a. SECURITY CLASSIFICATION AUTHORITY
N/A

2b. DECLASSIFICATION / DOUWNGRADING SCHEDULE
N/A

3. DISTRIBUTION / AVAILABILITY OF REPORT
N/A

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
ASQBG-A-89-034

5. MONITORING ORGANIZATION REPORT NUMBER(S)
N/A

6a. NAME OF PERFORMING ORGANIZATION
AIRMICS

6b. OFFICE SYMBOL (if applicable)
ASQBG - A

7a. NAME OF MONITORING ORGANIZATION
N/A

6c. ADDRESS (City, State, and ZIP Code)
115 O'Keefe Bldg.,
Georgia Institute of Technology
Atlanta, GA 30332-0800

7b. ADDRESS (City, State, and Zip Code)
N/A

8a. NAME OF FUNDING/SPONSORING ORGANIZATION
AIRMICS

8b. OFFICE SYMBOL (if applicable)
ASQBG - A

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

8c. ADDRESS (City, State, and ZIP Code)
115 O'Keefe Bldg.,
Georgia Institute of Technology
Atlanta, GA 30332-0800

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62783A	DY10	05-04-03	

11. TITLE (Include Security Classification)
Expert System Evaluation Methodology (UNCLASSIFIED)

12. PERSONAL AUTHOR(S)
Dr. J. W. Gowens

13a. TYPE OF REPORT

13b. TIME COVERED
FROM ______ TO ______

14. DATE OF REPORT (Year, Month, Day)
July 28, 1989

15. PAGE COUNT
40

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
Expert Systems, Artificial Intelligence, Knowledge Engineering, Expert System Sheels, Cost Benefit Analysis,

19. ABSTRACT (Continue on reverse if necessary and identity by block number)

The purpose of this guide is to provide procedures for quality assurance, impact assessment, cost benefit analysis and user acceptance testing.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED / UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Michael J. Evans

22b. TELEPHONE (Include Area Code)
(404) 894-3107

22c. OFFICE SYMBOL
ASQBG - A

DD FORM 1473, 84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

This research was sponsored by the Army Institute for Research in Management Information, Communications, and Computer Sciences (AIRMICS), the RDTE organization of the U.S. Army Information Systems Engineering Command (USAISEC). This effort was performed under Contract DAKF11-88-D-0011. This research report is not to be construed as an official Army position, unless so designated by other authorized documents. Material included herein is approved for public release, distribution unlimited. Not protected by copyright laws.

THIS REPORT HAS BEEN REVIEWED AND IS APPROVED

Accession For	
NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

s/ James Gantt

James Gantt, Chief
Management Information
Systems Division

s/ John R. Mitchell

John R. Mitchell
Director
AIRMICS

Table of Contents

Chapter 1

Introduction and Background

1.1 Purpose and Scope.

1.1.1 Purpose. The purpose of this guide is to provide the procedures to develop expert systems (ES) that can be used locally to assist in mission accomplishment and to aid decision processes of the local command. Using the guidelines given in this document, evaluation of expert systems developed using the Development Methodology, Volume 1, can be accomplished. This volume gives approaches to quality assurance, impact assessment, cost benefit analysis and user acceptance testing.

1.1.2 Scope. The guide has been written with the typical user assumed to be an officer or enlisted person with experience in microcomputers and some formal training in computer science, either in Army schools or a civilian educational institution. It can be used by commanders and/or staff personnel at all levels of command who wish to develop an ES to assist in the decision making process.

1.2 Background.

Expert systems are integrating with other software technologies. Operational expert system are replacing demonstrations and prototypes as the center of interest to organizations. These systems hold great promise for supporting decision support, especially in the military environment where decision time is short and accuracy paramount. As expert systems enter the mainstream Information Mission Area, they must be subject to the same evaluation criteria as traditional systems. This volume will focus in providing techniques that can be used to evaluate expert systems from conception through delivery.

So far, expert systems have concentrated on prototypes and demonstrations of the technology. In addition, most systems have been developed in an evolutionary or iterative methodology which focused on preparing the prototype with very little documentation beyond the source code. These environmental factors implied quality assurance, validation and verification, cost benefit and impact analysis have not been part of the development methodology.

1.2.1 Problems with Evaluation. Evaluation of computer systems has long been recognized as a very difficult problem. Issues that have been considered are integrity of data, reliability of the software, physical security, data security, verification of functionality, and audit trails. As expert systems mature, they must be subject to these same evaluation criteria, but the problems are greater for expert systems than for any information technology development from traditional computer systems.

There have been numerous approaches to evaluation of traditional computer systems: structured walk through, code inspections, and software testing. Some can be applied to expert systems and some cannot. Techniques such as structured walk through are not appropriate because expert systems do not have a priori paths defined through them that can be followed. There are numerous paths that can be followed based on the input data. They are data driven programs. Expert systems that are rule based have numerous rules, each rule having a life of its own; that is it is an independent function. It is a unique set of conditions that will cause it to fire when all conditions are met. This rule either puts facts into the data base or gives conclusions. If it puts facts into the data base, it will cause other rules to fire: i.e. forward chaining. With just a few rules, there are thousands of paths through the program.

1.2.2 Evaluation as Ongoing Process. Evaluation of the expert system development is an ongoing process throughout the development process: quality assurance, cost benefit, impact assessment, knowledge base validation, and acceptance testing. The system must be evaluated to determine the value or benefit relative to its cost early in the decision making process to determine if the project is viable. The alternative candidates must be evaluated in order to select one for development; since resources are scarce, one must select from among alternatives. As the system is developed it must be evaluated for knowledge base accuracy, integrity and reliability. Finally, the system must be evaluated by users in the context meeting the original functional definition defined for the system.

Expert system developments, just as their traditional counterparts, are generally conducted only as a single replication, hence rigorous statistical sampling methods will generally be unwarranted. This does not preclude good evaluation and this volume discusses evaluation methods that are appropriate.

1.3 Definitions.

The following were used as working definitions of terms and techniques in the preparation of the guide.

1.3.1 Quality Assurance. Quality assurance techniques are much the same as in any software development project. The techniques used in traditional software development are equally applicable in expert systems development. These include:

* Application of standards, such as this methodology.
* Definition and delivery of end products.
* Traceability, checking that decisions are documented, that a change is reflected in the software, that errors found in testing are corrected, and etc.

* Use of checklists to insure all steps in the process are properly conducted.

Paragraph 1.5 below provides the Quality Assurance points in the Expert Systems Development Methodology as specified in Volume 1 of this guide.

1.3.2 Impact Assessment. Impact assessment is the collection of techniques that seek to determine the organizational changes which will be required by the implementation of an expert system. It is a study in technological forecasting, both to determine feasibility and functionality.

1.3.3 Cost Benefit Analysis. Cost benefits analysis is the collection of techniques that are used to determine the life cycle costs of the expert system and compare them to the benefits derived from the system. These methods require a technological forecast from the impact assessment and other places.

1.3.4 Knowledge Base Validation. Knowledge base validation is the process of confirming the system emulates the expert behavior specified in the Functional Description and the Requirements Definition Document. The validation is conducted by experts using various techniques. Since this stage will include references back to the initial documentation of the system, this process includes some of the techniques termed verification in traditional software systems development. Compatibility with other systems and internal consistency are prime ingredients of this stage.

1.3.5 User Evaluation and Acceptance Testing. User evaluation and acceptance testing is the process of testing the software by the user community on actual problems in the working environment. The user community will seek to confirm the system meets all user expectations and outputs from the system match that specified in the Functional Description and Requirements Definition Document.

1.4 Format of the Guide.

Volume 2 contains four chapters which are a collection of evaluation techniques that may be used in all phases of the development of expert systems. Evaluation begins with the selection of the project. Simple selection criteria were given on candidate selection in Chapter 4 of Volume 1. In Volume 2, more penetrating analysis is given for evaluation of the project from initiation through acceptance testing.

Chapter 1, Evaluation Overview, gives an outline of the volume and motivation for the volume.

Chapter 2, Impact Assessment, is a guide for conducting an impact assessment of potential applications that were selected using the criteria in Chapter 4 in Volume 1.

Chapter 3, Cost Benefit Analysis, provides guidance for preparation of a cost benefit analysis for the expert system. This is an important step in the decision to build an expert system.

Chapter 4, Knowledge Base Validation, presents procedures for the validation and verification of the knowledge base in the expert system.

1.5 Quality Assurance.

Figure 1-1 shows some of the management and quality assurance actions that are required during the development cycle. Management is required to perform both formal and informal reviews of the system as it's being constructed. Informal reviews include the review of project objectives, approval of project scope, possibly also a project budget at this point. During the prototyping phase management should also informally review each prototype to make

sure it conforms to the objectives of the project. Formal reviews of the project include a review and approval of the recommended solution which is generally in the form of a requirements definition or solution definition document. To review and approve the implementation plan the training plan and knowledge based validation plan and finally to review and approve the final installation report.

Quality assurance actions are also important throughout the project. Informal quality assurance review is required for security requirements on the system, and to review the implementation, in particular to determine whether or not the users are satisfied with the system. Formal reviews for quality assurance include a review of the project estimates and the project plan to review and approve the recommended solution along with management of the organization. The combination of management and quality assurance actions insures a smoothly running project.

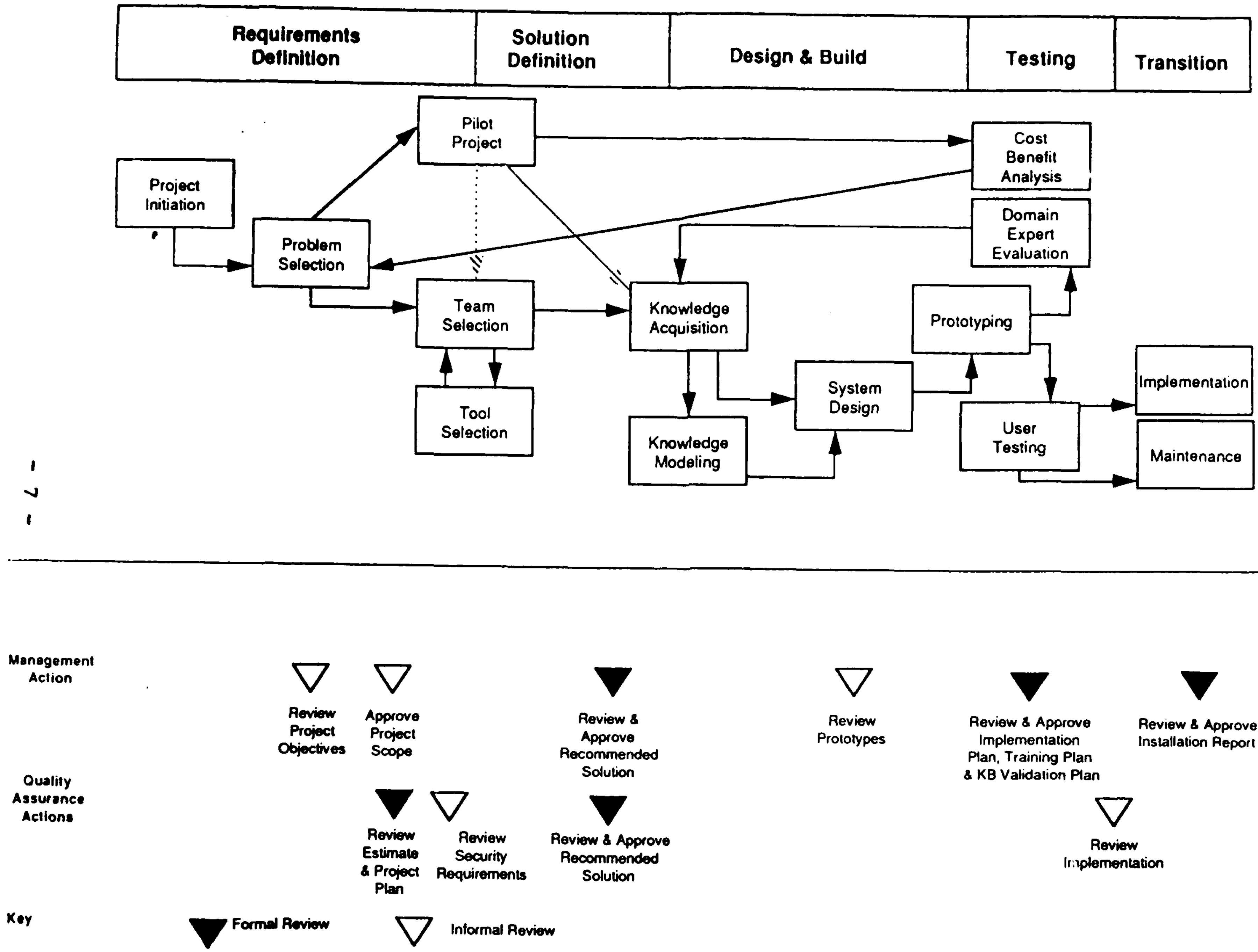

Figure 1-1. Management and Quality Assurance Actions

Chapter 2

Impact Assessment

2.1 Introduction.

The purpose of this chapter is to treat the analysis of the impacts of expert systems on organizations. First, there will be a generic discussion of impact assessment to orient the reader to the methodology of impact assessment together with its strengths and weaknesses. Then the specific case of the impacts of the use of expert systems on Army organizations will be discussed to specialize the general approach to the case of the military organization without focusing on any particular organization.

Impact assessment is a class of policy studies that systematically examine the effects on society that may occur when a technology, project, or program is introduced or modified. As well as immediate consequences, it considers those that are unintended, indirect, or delayed.

The development of impact assessment methodology goes back to the late 1960's and early 1970's when the passage of the National Environmental Policy Act and the development of the Office of Technology Assessment in the United States and serious interest by the OECD and many of its European members in the impacts of new technologies and projects led to the performance of impact assessments. Initially the desire to understand the impacts of developments led to many ad hoc approaches. However during the 1970's, the field began to stabilize, become more systematic, and allowed cumulative learning to occur.

Business and industrial organizations as well as governmental agencies have been involved in the practice of impact assessment.

Every impact assessment should be both valid and useful. Since impact assessment deals with the future which is as yet unknown, the congruence between the projected and the actual impacts is unknown. Validity can be approached by developing a sound cause-effect understanding, by treating the analysis even-handedly both in coverage of major issues and representation of the diverse perspectives of the parties involved. The utility of an assessment is determined by its relevance, timeliness, credibility, and communicability.

2.2 The Structure of an Impact Assessment.

This section deals with the components of an impact assessment. Together they possess a logical, but not necessarily temporal, order. Grouped together they describe the flow of an impact assessment, and at the same time the actual process of socio-technical change. A technology, project, or program interacts with its societal context to produce impacts. These impacts are modified by actions taken by institutions or possibly individuals. These modified impacts alter both the technology, project, or program as well as its societal context. The process continues through this feedback in a sort of spiral evolution.

There are ten components of an impact assessment:

1. Problem definition and bounding
2. Description of the technology, project, or program to be assessed
3. Forecast of the development of the technology, project, or program to be assessed
4. Description of the context with which the technology, project, or program will interact
5. Forecast of the context with which the technology, project, or program will interact
6. Impact Identification
7. Impact Analysis
8. Impact Evaluation
9. Policy Analysis
10. Communication of Results

These components are not all required in every assessment, nor do they necessarily follow a fixed, linear order in every impact analysis. Rather they serve as a checklist of important functions that should be dealt with in the assessment process. However, there is a logic in the order of their presentation that is useful to understand. In reality, components may be omitted, done in different orders, be enmeshed in various feedback loops, and otherwise be ordered to meet the needs of the assessment at hand. The components are discussed individually below.

2.2.1 Problem Definition and Bounding. This is the first component in any impact assessment. Here the assumptions involved in the assessment are articulated and the boundaries of the investigation specified. The nature and scope of the study are specified. The breadth and depth of coverage are determined based on objectives and available resources. An important part of this component is the identification of the parties at interest in the assessment. These are the persons or groups affected by the subject of the assessment and those that may make decisions involving it.

Bounding an assessment is an ongoing activity. While bounding substantially occurs at the beginning of a study, information is often uncovered that requires some shifting in study scope and emphasis as the assessment develops. This flexibility allows unanticipated findings to be dealt with readily in the course of the assessment. There is a definite advantage of dealing with assessment as an ongoing program, rather than a single shot study, in the case of a technology or development that is evolving. Here, changes in either the assessment subject or its implementation may substantially alter its impacts. An ongoing program of assessment allows changing conditions to be assessed and impact projections to be altered as the situation matures.

There are a number of important areas for bounding. It is important to determine the time horizons of the study. Impacts change over time, and impacts that are insignificant at one time may become substantial at a later time. Witness the difference in automotive air pollution in Los Angeles in 1915 and 1975. The spatial and/or organizational extent of the study is also important. Are impacts on the world being considered, or just the United States? Is the study dealing with impacts on a single firm or an entire industry? The technology and its range of applications or the extent of the project or program to be assessed need to be determined so that the study is focused on a clearly delineated object. It is also important to specify generically the range of impact sectors and policy options to be considered in the study.

A very useful technique to guide the bounding of an assessment is the bounding microassessment. This is a very quick and rough run-through of the entire assessment at the beginning of the study to provide the assessors with some idea of the potential range of the full-scale assessment. The microassessment can be conducted by a single person or a small group at a level of effort beneath one person-month. It should be broad ranging and superficial to quickly identify the major impact areas and policy considerations. It generally relies on the available literature and the views of knowledgeable persons. Typically the microassessment will generate more than enough material to intelligently bound the assessment as a whole.

2.2.2 Description of the Technology. This component of the study answers the question: "What is to be assessed?" in detail. The level of detail depends on the depth of the study. Critical in how this component is dealt with is the status of the emergence of the assessment subject. That is, something that is well known and mature allows a much fuller description than a technology or

project that is still in the conceptual stage. Principal sources of information for this descriptive phase are the literature and subject matter experts. The description phase is largely an empirical data gathering activity. The development of a framework, matrix, or checklist from which to structure the description is often very useful.

2.2.3. Forecast of the Development. Because of the future nature of assessment, it is necessary to forecast the development of the assessment and to track the changes in impacts over time. There is substantial evidence that changes in many important technological parameters have been orderly implying that time series analysis can be used to track them. Technological changes respond to need as determined by the allocation of resources for innovation and by social control through regulation. Tracking these factors allows for the anticipation of some forms of technological change.

Five major families of forecasting techniques can be mentioned briefly to orient the reader to the possibilities of forecasting in different situations. In general, a forecast incorporating a number of approaches gives more confidence in limiting uncertainty about the future than one based on a single approach.

2.2.3.1 Monitoring. Monitoring is not strictly a forecasting technique. However it is treated as one because of its importance in gathering information for the forecasting process. It is both the simplest and most widely used technique. It consists of scanning the environment, usually the literature, for information relevant to the topic being forecast. It is most important that selectivity be exercised in obtaining information as masses of undigested information typically remain just that. Finally, it is important that the information be structured for use in forecasting.

2.2.3.2 Expert Opinion. Expert opinion techniques allow the views of experts to be effectively tapped for forecasting. Perhaps the most widely used technique is Delphi. This technique allows for both anonymity of the individual expert and feedback of additional information to the experts. The experts are first identified. They are given a survey, carefully pretested to eliminate ambiguities, that elicits their views relative to the forecast. The survey is then analyzed, and the results fed back to the participants. They are then asked to answer the survey again in the light of the responses. They may be asked to explain the reasons for their result, if it is on the extremes of the responses. The process is iterated until the responses become stable. Typically they will converge. However, this may be about more than one point. Three or four rounds are usually necessary for convergence.

2.2.3.3 Trend Analysis. Trend analysis is viewed by some forecasters to be nearly the sum total of forecasting. It has been widely used in various economic and demographic forecasts. Trend analysis relies on accurate time series data which is then extrapolated into the future. There are many methods of trend extrapolation, ranging from merely "eyeballing" the data to the much more sophisticated Box-Jenkins and ARIMA approaches.

2.2.3.4 Modelling. Modelling is a generic approach subsuming many individual techniques. The object is to simplify a part of the world in such a way that its main features are grasped. The development of the model over time produces the forecast. In many techniques, such as interpretive structural modelling, human judgment is incorporated into the modelling process as a critical component. While physical models and "boxes and arrows" models may be used effectively, many models are computerized such as feedback dynamics models using the DYNAMO simulation language. The most

critical aspect of successful modelling is the quality of the assumptions underlying the model.

2.2.3.5 Scenario Construction. Scenario construction is the final technique. Scenarios are snapshots of a particular aspect of the world at some time in the future. In addition, the term has been used as well to apply to the developmental path from the present to the future state. The latter may also be referred to as future histories. Scenarios are imaginative devices that may incorporate media and literary techniques for presentation. As such, they are effective devices for communicating technical material to non-technical audiences. Because a scenario can incorporate both quantitative and qualitative information, it is an effective integrating device for information inputs from various sources.

2.2.3.6 Strengths and Weaknesses. All of these families of techniques have their strengths and weaknesses. Monitoring is quite useful except when the development is so far in the future that there is no sound information available. Expert opinion depends on the existence and involvement of experts. Trend extrapolation depends on the existence of good quality time series data. Modelling rests on the ability to be able to make sound assumptions from which a model can be constructed. Scenario construction is very helpful for integration and communication, but it may degenerate into imaginative fantasy if no relevant and sound information exists.

2.2.4 Description of Technology Context. Societal context description emphasizes that part of the world with which the assessment subject will interact intimately. One effective approach is to use the concept of the technological delivery system (TDS). The TDS is a model of the institutions and resources that are involved with the assessment subject together with their interactions. The institutions are political, economic, social,

legal, etc. The TDS can be exhibited as a "boxes and arrows" diagram that indicates both the institutions involved and their interactions. Relevant information can be incorporated within this framework.

2.2.5 Forecast of New Technology Context. In this component, the societal context is forecast independent of the involvement of the assessment subject. Thus, this is not the same as social impact assessment because no impacts are being considered. This is one of the least used and most difficult of the assessment components. Forecasting techniques are described in Paragraph 2.2.3. above. The techniques for economic and demographic forecasting are quantitative and widely used. However other aspects of the societal context lack any clear track record of forecasting. Techniques used for this are largely qualitative with scenarios playing a major role.

2.2.6 Impact Identification. Impacts are the products of the interaction between the assessment subject and its context. In this step, the principal areas of impact are identified using scanning or tracing techniques. Typically the principal areas of potential impact are identified and then these areas are subdivided using a checklist or its two dimensional version, the matrix. This scanning procedure relies heavily on the judgment of the assessors to identify the impact areas deemed sufficiently significant for further analysis. In tracing, immediate impacts are identified by judgment and causal chains are constructed to link the first order impacts to higher order impacts.

2.2.7 Impact Analysis. The impacts identified are subject to detailed analysis to determine their character, magnitude, and likelihood. The methods of analysis used are specific to the area of impact being analyzed. The acronym EPISTLE has been used to divide the impacts into the major categories of environmental,

psychological, institutional, political, social, technological, legal, and economic. The methods used in analyzing these impacts represent the state of the art of available analytic techniques in such disciplines as economics, atmospheric science, ecology, hydrology, organization theory, law, and sociology. This component requires substantial area expertise to perform effectively. There is a danger that subject matter experts will spend almost all their time on components 2 and 7 to the exclusion of everything else.

2.2.8 Impact Evaluation. Impacts, once identified and analyzed, must be evaluated in the light of the assumptions and boundary conditions guiding the assessment. The costs and benefits resulting from the assessment subject are assessed, and this information forms the basis for policy formulation. It is important to remember that different parties at interest evaluate the same impacts from different perspectives. What is a desirable impact to one party may be a disaster to another. It is thus important to clearly define the evaluation criteria and their contextual application. It is also important to note alternate stakeholder value sets that should be understood in the process of evaluation.

2.2.9 Policy Analysis. Policy analysis is the answer to the question: "Given that the impacts have been identified, analyzed, and evaluated, what can be done to deal with them, when, by whom, and with what results?" A second question that may be asked in some cases is: "What should be done about these impacts?" Usually policy considerations should begin in the bounding phase of the assessment. Assessors should be sensitive to policy implications throughout every component of the study. It is good practice to get input on potential policy actions and their impacts from the major actor groups -decision makers and impacted publics- the policy community. The question of making specific recommendations is one that must be asked and answered individually for each study.

In some cases, the sponsors of the study merely want informational input for decision making or options without recommendations; in others, explicit recommendations for action are expected and indeed necessary. It is important to remember that the assessors structure the analysis, and, insofar as decision makers accept this structure, they will find themselves within the analysts framework for laying out optional courses of action.

2.2.10 Communication of Results. Impact analyses are performed to be used to affect decisions in the real world. Thus, it is critical that they be organized in a form that they can be accessible to the user group. In particular, highly technical analyses are of little interest to non-technical users. Thus, the results of technical analyses need to be expressed in whatever language and/or media are accessible to the decision makers and other parties at interest. It is the responsibility of the assessors to "package" the analytical results appropriately for use. It has also proven to be quite effective to communicate the study's progress to the users throughout the course of the study so that they will be able to ask questions, offer criticism, and input new information during the course of the study. This makes the users take a proprietary interest in the study and, if the interaction is well handled, increases the credibility of the assessment for them.

2.3 The Impacts on Military Organizations.

In this section the components of an impact assessment, described above, are specialized to the case of the impacts of expert systems on Army organizations. This narrowed focus, however, leaves sufficient latitude to deal with substantial variations in expert systems, Army organizations, and operating environments.

In bounding the study, the interests of the organization in performing the study, including the resources and time available

for the study, need to be considered. In particular, the values of the organization as expressed in its mission statement need to be made as explicit as possible to establish the frame of reference within which the study is to be conducted. The expectations for the study need to be made clear at the beginning, subject to the proviso that they may be revised during the course of the study.

The description of the expert system should stress its functional capabilities and operating requirements, both in terms of hardware and personnel. The limitations of the system as well as its strengths should be made clear.

Forecasting the evolution of the expert system should stress, as above, functional capabilities and operating requirements. Potential paths of incremental improvement should be explored as well as the possibility for major, discontinuous change in the entire operating basis of the system. The impacts of the latter will be more striking than those of the former. Where time varying parameters can be established, time series analysis may be used. Expert opinion may also be useful in areas where explicit, quantifiable parameters are not clear. Scenarios constitute a useful integrating tool.

Context description refers primarily to those areas of the organization that will interact with the expert system, the overall environment of the organization including mission statement, a "champion" for the introduction of the expert system, and factors external to the organization that may influence the use of expert systems such as conditions in the Army, Congressional constraints and the overall budgetary environment.

Context forecasting focuses on the evolution of those factors of the context changing substantially without factoring in the use of expert systems. It is important to remember that this deals with

the development of the organization, trends in Congress and changes in the national economic and political situation. This component is not the same as social impact analysis.

Areas to be considered in impact identification include selection and training of personnel, job responsibilities of personnel, management patterns, distribution of costs and benefits, and changes in mission programming. In general, people related issues within the organization, organizational structure and mission need to be seriously considered.

Techniques to be used in impact analysis may be drawn from such areas as industrial psychology, organizational behavior and organizational sociology, and cost-benefit analysis. Emphasis should be placed on changes in functional responsibilities, skill requirements, intra-organizational communications, training, and management structure. Analyses should determine changes in organization performance and additions of capabilities both with regard to existing products and the development of new products and product lines depending on the type of expert systems implemented.

Emphasis in impact evaluation should be focused on the effectiveness and efficiency of the impacts. Overall, the utility of the impacts to the organization should be considered. Policy analysis should be designed to integrate the expert systems into the organization with a minimum of disruption and a maximization of utility to the organization. Options may be in the areas of functional organization of work, responsibilities of personnel, training and selection of personnel, changes in product line, etc.

Communications should be maintained throughout the study with the key persons in the organization who are responsible for dealing with expert system related issues. The study output should be put in a form that is useful to the decision makers within the

organization. By "staying close" throughout the study, there should be no surprises and the results of the study will be sold to the decision makers throughout the assessment process.

2.4 Concluding Observations.

In considering the impacts of expert systems on Army organizations, it may be helpful to think of other systems previously introduced to increase the productivity and effectiveness of Army organizations. Analogies with these systems and their impacts may be a fruitful source of information in guiding the analysis of the impacts of expert systems.

It is most critical to understand that expert systems are not a panacea. Nor do they replace human intelligence in critical applications. Their strengths are to support, aid, and sustain the human decision making process.

Finally, impacts have a tendency to interact with other impacts to produce unanticipated consequences. In addition, policy options can also couple with other options to produce complex effects. The very clear message in impact analysis is that a systemic treatment is necessary. Attempts to treat impacts in isolation and to deal with them without considering other parts of the technology delivery system can lead to undesirable outcomes. Impact assessment is systemic, cross-disciplinary, and cross-organizational.

Chapter 3

Cost Benefit Analysis

3.1 Introduction.

It is the purpose of this chapter to provide guidance on preparing cost benefit models for the evaluation of candidate expert system projects. The approach will be build an estimated cost model and compare it to an estimated benefit model. The comparison of these two models will be by subjective analysis of the decision maker. Be careful not to make the cost/benefit models too detailed because they will become unmanageable. Yet they should not be too general because they will be too insensitive to changes in significant system parameters.

3.2 Matrix Model.

The foundation of the cost model is the arrangement of the cost elements into a matrix format as shown in Figure 3-1. The organization of the matrix provides a convenient means of systematically analyzing and synthesizing the cost elements. The matrix also fits easily into a Lotus-123 worksheet model to allow rapid calculation of the numeric values when performing "what if" modelling under different assumptions for the system.

An inherent advantage in preparing the cost matrix for use in the cost/benefit analysis is it forms a checklist of all the cost components and physical functions to be considered. Thus, the matrix helps in pointing out possible omissions in the total cost function.

A benefits matrix similar to the cost matrix can also be prepared. The benefits matrix provides the same advantages as the cost matrix and will ultimately be used in a comparison to determine if the

Matrix Model

Phase	Physical Functions	Cost Functions			
	Modules	Equip	Maint	Admin	Totals
R&D	1	$	$	$	sum $
	.				
	.				
	N	$	$	$	sum $
Total R&D		$	$	$	sum $
Invest-ment	1	$	$	$	sum $
	.				
	.				
	N	$	$	$	sum $
Total Invest.		$	$	$	sum $
Annual Operat.	1	$	$	$	sum $
	.				
	.				
	N	$	$	$	sum $
Total Operat.		$	$	$	sum $
Oper.	Total operation costs * No yrs	$	$	$	sum $
R&D	Total R&D + Invstmt	$	$	$	sum $
All	Total sys Cost	$	$	$	sum $

Figure 3-1. Cost (Benefit) Matrix

proposed expert system is viable. Net value to the organizationis computed by taking the difference between the total costs and the total benefits.

3.2 Determination of Matrix Elements.

The next step is the computation of the various matrix elements for both the cost side as well as the benefit side of the equation. Determination of the specific elements of cost is necessary in each project individually, but they usually consist of equipment, software, research activities, administration, operation of system, training, and etc.

3.2.1 Determination of Costs.

3.2.1.1 Historical costs. The most straightforward cost determination is to determine the historical costs for similar functions. History is not always available, especially in expert systems development, but there are some historical costs that can be found.

3.2.1.2 Surrogate Costs. The cost of an element might be estimated by estimated a related element then computing the desired cost function through a formula. For example, the cost of programming might be associated with the cost of the computer. This relationship would then be used to estimate programming costs based on the cost of the computer. The use of surrogate cost functions is not as accurate as historical costs but it is sometimes the only thing available.

3.2.1.3 Aggregation of Components. The aggregation of components of the costs elements is another approach to estimating costs. For example, Module 1 R&D might consist of engineering in human factors, reliability, maintainability, mechanical, electronic, communications, and etc.

Similarly, the other cost functions can be derived through aggregation of their component parts.

3.2.2 Determination of Benefits.
The estimation of benefits follows the same line as the costs. One real problem in estimating benefits, however, is that they are often not measurable directly. Therefore, surrogates are much more often used for benefits estimation than for costs. Benefits in the Army environment are primarily related to cost savings on a targeted task or set of tasks, since the Army does not generate revenue.

These cost savings generated by the expert system will very difficult to estimate if no impact assessment has been conducted, because it is in the impact assessment that the technological forecast is made. Thus, even if no impact assessment is done formally, it must be done informally to forecast the changes in user behavior in the organization as a result of implementing the system.

3.2.2.1 The Value Approach to Benefit Estimation. One approach to determining benefits is to estimate the "value" of the system implementation to the user organization. This approach focuses on "value" as the "revenue" of the implementation. The value is given in dollars for later computations, but the amount assigned to the use of the system by the organization is determined by an estimate from a panel of users. To use this approach, a pilot project (see Volume 1, Chapter 3) will usually be required. If no pilot project is conducted, then at least a detailed functional description is necessary to allow a panel of users to estimate the value of the system.

Value analysis focuses on:

* identifying significant intangible benefits for the specific expert system,
* quantifying these benefits in value terms and translating them to the same unit of measure as the cost terms, normally dollars, and
* establishing a decision rule for identifying the significance of the proposed system.

Of course, in computing the total benefits, any tangible benefit elements would also be utilized.

The evaluation of the proposed expert system is determined by asking a selected panel of potential users to analyze the expert system in terms of the attributes of the system. The tradeoffs inherent in this evaluation will be determined by a four phase process:

* Phase 1: The identification of appropriate expert system benefits using the Delphi (see Chapter 2) technique and a literature search.
* Phase 2: The grouping of the individual expert system benefits into groups of related benefits.
* Phase 3: Quantification of the benefits by assigning dollar values to the benefits within each group.
* Phase 4: Establishing a decision rule for determining the value of the system to be compared to the cost matrix.

3.2.2.2 Phase 1. The panel of users is convened and asked to evaluate the expert system and to create a list of benefit attributes. This phase is conducted by a moderator who is impartial to the development. The users are asked to individually

evaluate the system and the benefits it might have. The moderator collects the lists and creates a master list containing all the benefits given by the group. The list is then circulated back to the group for further consideration. The process is repeated several times until consensus is reached on the benefits list. Examples of possible benefits are:

* Clerical time and labor saving
* Better utilization of data
* Improves communication between managers
* Improves planning and control
* Improves utilization of management time
* Deeper and wider exploration of alternatives
* Improves decision making capability
* Clearer appreciation or understanding of problem

3.2.2.3 Phase 2. The benefits are then grouped into clusters that represent various views of the benefits. Three potential areas for benefits to be grouped into are: operational, managerial or organizational, and personal. after the benefits are divided among the groups, a null benefit is added to each group. This represents no change in the organization's operations. This null benefit will be used in all the following computations along with the benefits of the system. The null benefits provides a benchmark for the value associated with no change. Presumably the system will have benefits which are all more than the null. These benefit groups are then used in Phase 3 to determine their value.

3.2.2.4 Phase 3. In Phase 3, values are assigned to the benefits derived in the previous phases. The value is assigned by first creating a rank ordering on the benefits within each group and a rank ordering of the groups as a whole. The rank ordering is computed by the moderator by asking each panel member to independently rank each benefit within each group and to rank each

group. The average ranking of benefits within and between each groups is then computed by the moderator to get the rankings. The rank ordering provides a measure of the relative importance of the benefits, but no insights into the strength of importance of each benefit.

The strength of each benefit is obtained by asking each panel member to independently assign a dollar value to each benefit that represents its annual value to the using organization. This dollar value is each member's perception of "annual cost saving" or "annual revenue" for the benefit. The dollar benefit is assigned to each benefit individually as well as to the group as a whole. The reason for evaluating each benefit individually and as a group is that sometimes the value of a group of benefits is more than the sum of its component parts. This effect is termed the synergistic effect: the whole is greater than the sum of its parts. In order to get convergence on these values, there will usually need to be two or three rounds of estimation by the panel with feedback between each round about the value of the panel estimates.

The values from each panel member are his subjective estimates based on his experience in the using environment of the proposed system. The null benefit should be included in this assignment of dollar value. The values are again averaged by the moderator to obtain a group dollar value. The total of the benefits obtained at the conclusion of the process is the "value" of the system. These annual figures can then be used over the life of the system to estimate total life cycle benefits.

3.2.2.5 Phase 4. The decision facing managers in justifying the continued development of the expert system is whether the proposed system has a value which exceeds the cost of development. The cost matrix total life cycle costs are compared to the total life cycle values. When value exceeds costs by a large margin, the decision

to continue to develop the system is generally straight forward. If they are very close together, then a management judgment must be made on the continuation of the system. There may be overriding mission requirements that imply continuation. Other conditions, such as technical risk, may weight the decision against continuing.

Chapter 4

Knowledge Base Validation

4.1 Introduction and Background.

4.1.1 Introduction. Validation and verification (V&V) have been the subject of numerous debates about whether a software system can be truly validated. This chapter will not delve into the philosophical problems of validation but will simply attack the problem as if it is soluble. The chapter will provide the best known validation procedures currently in practice in expert systems development. The argument as to whether they are optimal will be left to other arenas for discussion. Validation of the knowledge base is the topic of this paragraph while the verification (see definition below) will be discussed in the following paragraph.

4.1.2 Background. Validation is the process of determining whether the system "correct" in the sense it meets an acceptable level of accuracy in its decision making relative to a standard. Thus, **validation** substantiates whether the **right system has been built.** Verification is the process of determining if the completed system meets the Requirements Definition Documentation (see Volume 1, Chapter 11) and has correctly implemented its specifications. Thus, **verification** substantiates whether the **system has been built right.**

4.2 Validation Types.

Validation techniques fall into four categories: internal or logical consistency, historical validation, predictive validation and multistage validation. The last category is staged application of the first four categories of validation. The following paragraphs discuss each of the types of validation.

4.2.1 Internal Consistency. The validation of internal consistency or logical consistency is the process of analyzing the program written in the expert system shell for errors in logic, omission of rules and redundancy of rules. Some appropriate checks are given below. These checks will normally be made by the programmer manually; however, there are several projects underway that will eventually make these checks automatically for some expert system shells.

4.2.1.1 Structure Checks. Structure checking involves the checking of the knowledge base for dead-end and unreachable nodes, redundant rules, and rule cycles. A dead-end occurs when a clause if the left-hand side (LHS) of a rule does not match any fact, goal or frame in the knowledge base. An unreachable node is a fact, goal or frame in the knowledge base that cannot be matched by a rule clause.

A rule is redundant if it is duplicated by or included into another rule. Two rules duplicate one another if they have the same LHS and right-hand side (RHS) clauses, possibly in a different order or with different variable names. For example:

```
Rule A:
     IF   subject is male
          and subject is parent
     THEN
          ASSERT subject is father

Rule B:
     IF   person is parent
          and person is male
     THEN
          ASSERT person is father
```

Notice the order is different and the two rules use different variables.

A rule is included in another if its LHS is a subset of another LHS or its RHS is a subset of another RHS or both conditions exist.

For example:

```
Rule A:
     IF   person is tenured
          and person is not staff
     THEN
          ASSERT person is university member

Rule B:
     IF person is tenured
     THEN
          ASSERT person is university member
```

The last check for structure is the rule cycling check. A rule forms a direct cycle when the same clause exists on both the LHS and the RHS of a rule. This may cause the rule to fire repeatedly creating an infinite loop. An indirect cycle is formed when there are two or more rules involved in the cycle.

4.2.1.2 Logic Checks. Logic checks determine whether the rule base is consistent and whether any clauses are irrelevant. Two rules with duplicate LHS's are directly inconsistent if they contain the same RHS's except that in one the RHS is affirmed and in the other the RHS is denied. An indirect consistency occurs when three or more rules are involved in the inconsistency.

Relevance is checked by determining if two rule clauses have duplicate LHS's except that in one LHS a clause is affirmed and in the other rule the clause is denied.

4.2.1.3 Semantic Consistency. Semantic consistency is a check to determine if logical world knowledge has been violated. For example: an Army member cannot be both on active status and inactive status at the same time. This is not consistent with world knowledge. Similarly if gaps are left in the knowledge base, especially in the algorithmic area, there would be a semantic inconsistency.

4.2.2 Historical Validation. Historical validation is the process of determining whether the expert system performs like an expert on cases that have already been completed by human experts. It is likely that some number of these cases were used in building the system. Presumably the system has been built to handle these cases. Therefore, any test cases used in system development should be discarded in the validation process.

For a fair validation, the test cases should be selected randomly using stratified statistical sampling techniques. For instance, if an expert classifies a diagnosis as A, B, or C and these diagnoses occur respectively 80 percent, 15 percent and 5 percent of the time, then 80 percent of the sample test cases should come from diagnosis A, 15 percent from diagnosis B, and 5 percent from diagnosis C. The number of test cases used in the validation will affect the confidence in the test results.

If no test cases are available or if all the cases were used in developing the system, then either the system will not be validated historically or the experts can create some random test cases. These random test cases are likely to have considerable bias in them and the value of such a validation is problematical.

The procedure of the validation is to compare the human expert's results with the expert system's results for the same case. The validation can include validation of intermediate results, final results, the reasoning of the system or all three. Validation of the reasoning process is particularly important to insure the system is obtaining the correct result for the correct reasons. Getting the correct result for the wrong reasons is not validation. Typically, we are concerned with validating the reasoning process early in the development of the prototypes (see Volume 1, Chapter 9) and with the final result when the knowledge base is nearer completion.

4.2.3 Predictive Validation. Predictive validation is essentially the same as historical validation, except that it is performed on newly initiated cases not completed or analyzed by the human experts. After the system has completed the case, the human expert performs the case analysis without knowledge of the results of the expert system. Obviously, predictive validation processes are conducted as the system nears delivery to the users. In fact, it might be termed beta testing. The system results are then compared to the human expert results.

4.2.4 Multistage Validation. Multistage validation is the combination of the three procedures given above. Most expert system development will use some sort of multistage validation procedure. The stages can be performed as many times as there are case data, developer patience and money for the project.

4.3 Verification.

Verification of the expert system entails providing a system to the using community along with the original documentation for the system. Users then exercise the system in their daily work environment. Several activities can take place in the verification.

4.3.1 Requirements Definition Document Review. Review of the Requirements Definition Document by the user groups is to determine if the system provides the functionality defined in its documentation. The review consists of exercising each of the functions to determine that they exit and operate.

4.3.2 Event Logging. Event logging is a technique that records the use of the system. Event logging can be a built-in feature of the prototype software or it can be a manual log. Event logging simply seeks to describe the activity on a specific case and its

result. In effect it is score keeping of the usability of the system as perceived by the user. During the event logging evaluation, any discrepancies noticed relative to the Requirements Definition Document should be noted in the log.

4.3.3 Critical Case Analysis. Critical case analysis is a technique of identifying cases that make a point dramatically or are particularly important to the organization then using them in the analysis. A clue to the existence of a critical case is the statement by a user that "if it doesn't make it here, it won't make it anywhere." Studying a few critical cases does not technically permit generalization to all possible cases but logical extensions can often be made from the penetrating analysis of a critical case. The results of the critical case analysis should be checked against the functional description, Requirements Definition Document and System Specification to determine that the system has been built as specified.

4.3.4 Work Profile Analysis. Work profile analysis is a technique that determines how the work activity of the domain expert may change as a result of the system installation. Work profile analysis requires a pretest interview to elicit the work profile of the user prior to installation of the system. There is a corresponding post-installation interview to determine the changes in the work profile, if any. The results of the survey should be checked against the functional description, Requirements Definition Document and System Specification to determine that the system has been built as specified.

Index